DOI:10.12113/202002005

基于全基因组序列的弯曲菌特征分析

孙 磊[1,2,3*],杨臻辉[1,2],恽 茜[3,4],黄金林[3,4]
(1.扬州大学 信息工程学院,江苏 扬州 225127;2.扬州大学 人工智能学院,江苏 扬州 225127;
3.江苏省人兽共患病学重点实验室,江苏 扬州 225009;
4.江苏省动物重要疫病与人兽共患病防控协同创新中心,江苏 扬州 225009

摘 要:空肠弯曲菌(*Campylobacter jejuni*)和结肠弯曲菌(*Campylobacter coli*)是引起人类腹泻的主要致病菌。传统生化方法在鉴定弯曲菌时存在步骤多、耗时长、通量低等问题。本研究通过利用生物信息学方法对弯曲菌全基因组进行序列、基因注释、耐药基因、多位点序列分型以及CRISPR-Cas系统等分析,挖掘能够有效区分空肠弯曲菌和结肠弯曲菌的高分辨力特征。实验结果表明,空肠弯曲菌和结肠弯曲菌在基因组序列长度、GC含量、基因数量、多位点序列分型以及CRISPR-Cas系统等方面存在显著差异。同时,研究还发现了一段在空肠弯曲菌基因组中广泛存在的高分辨力CRISPR重复序列。这些特征可用于构建能够准确鉴别空肠弯曲菌和结肠弯曲菌的生物信息学方法。

关键词:弯曲菌;全基因组;多位点序列分型;CRISPR-Cas系统
中图分类号:Q93 **文献标志码**:A **文章编号**:1672-5565(2020)04-254-09

Feature analysis of *Campylobacter* based on whole genome sequences

SUN Lei[1,2,3*], YANG Zhenhui[1,2], YUN Xi[3,4], HUANG Jinlin[3,4]
(1.*School of Information Engineering, Yangzhou University, Yangzhou* 225127, *Jiangsu, China*;
2.*School of Artificial Intelligence, Yangzhou University, Yangzhou* 225127, *Jiangsu, China*;
3.*Jiangsu Key Laboratory of Zoonosis, Yangzhou* 225009, *Jiangsu, China*;
4.*Jiangsu Co-Innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonosis, Yangzhou* 225009, *Jiangsu, China*)

Abstract: *Campylobacter jejuni* and *Campylobacter coli* are the main pathogenic bacteria that cause diarrhea in humans. Traditional biochemical methods have problems such as multiple steps, time consuming, and low throughput when identifying *Campylobacter*. In this study, several analyses including sequence analyses, gene annotation, drug resistance gene prediction, multilocus sequence typing (MLST), and CRISPR-Cas system finding were performed on the whole genomes of *Campylobacter* using various bioinformatics methods, thereby exploring highly effective features for distinguishing *Campylobacter jejuni* from *Campylobacter coli*. Experimental results manifest that *Campylobacter jejuni* and *Campylobacter coli* have significant differences in genome sequence length, GC content, gene number, MLST, and CRISPR-Cas systems. In addition, we found a highly distinguishable CRISPR repeat sequence that is widespread in *Campylobacter jejuni*. It is concluded that these features can be used to construct bioinformatics tools for identifying *Campylobacter jejuni* and *Campylobacter coli* more accurately.

Keywords: *Campylobacter*; Whole genome; Multilocus sequence typing; CRISPR-Cas system

弯曲菌(*Campylobacter spp*)是一类能够引起人类腹泻的人兽共患病病原菌[1]。因弯曲菌感染引起的肠炎病例数仅次于沙门氏菌和志贺氏菌[2]。在已发现的弯曲菌种类中,空肠弯曲菌和结肠弯曲菌是引起人类腹泻的主要致病菌,90%以上的病例由这两种病原引起[3],而其中空肠弯曲菌和结肠弯曲菌分别约占90%和10%[4]。空肠弯曲菌和结肠弯曲菌广泛存在于自然界,通过传统流行病学方法

收稿日期:2020-02-16;**修回日期**:2020-03-07.
基金项目:国家重点研发计划项目(No.2018YFD0500500),江苏省人兽共患病学重点实验室资助项目(No.R1805).
*通信作者:孙磊,男,副教授、硕士生导师,研究方向:生物信息学.E-mail: sunlei@yzu.edu.cn.

并不能准确确定传染源[5]。在 WHO 食品安全工作计划中,弯曲菌被划列为重点检测的食源性致病菌之一,许多国家也相继开展了对弯曲菌的监测。

目前弯曲菌检测主要针对空肠弯曲菌。传统检测方法包括前增菌、细菌分离和种的鉴定等步骤,其工作强度较大且耗时长,需要约 5~6 天才能完成检测[6]。同时,这些传统检测方法所依赖的关键生化反应—马尿酸盐水解试验可能导致假阴性和假阳性结果[7],从而影响后续分析。近年来一些基于常规或定量 PCR 的方法被用于弯曲菌的检测。荧光定量 PCR 方法较常规 PCR 具有更好的灵敏度,但仍存在不能有效区分细菌的死活状态以及不能获得活的培养细菌等问题[8-10]。同时,这些 PCR 方法所使用的试剂昂贵,且实验步骤多,容易产生样本间的交叉污染,从而导致假阳性和假阴性结果。

近年来全基因组测序技术(Whole-genome sequencing,WGS)开始被用于弯曲菌研究。测序数据在经处理和分析后可用来表征弯曲菌的不同种系,或用来快速识别群落的基因型特征,例如毒力和耐药性等[11]。同时,基于全基因组测序技术的弯曲菌研究也产生了大量数据(如存储于美国生物信息技术中心 NCBI 的弯曲菌全基因组序列等),而这些数据可用于进一步的信息挖掘。因此,本文将基于弯曲菌的全基因组序列,利用相关生物信息学方法,对弯曲菌的基因组序列、基因注释、耐药基因、多位点序列分型(Mutilocus sequence typing,MLST)和簇状规则间隔回文重复(Clustered regularly interspaced palindromic repeats, CRISPR)-Cas (CRISPR-associated)系统等进行分析,以挖掘出能够准确区分空肠弯曲菌和结肠弯曲菌的高分辨力特征,以帮助建立能够快速、准确地检测弯曲菌的生物信息学方法。

1　材料与方法

1.1　菌株的选择

本研究共包含 120 株空肠弯曲菌(空弯)和 22 株结肠弯曲菌(结弯),其菌株的全基因组序列下载自 NCBI(https://www.ncbi.nlm.nih.gov/,按关键字“Campylobacter complete genome”搜索,结果中仅下载全基因组核心菌株)。以下实验过程均在这 142 株弯曲菌上展开。

1.2　全基因组序列分析

基因组序列长度和 GC 含量是生物最基本的遗传特征[12-13]。本文通过 BioPython[14] 获得空肠弯曲菌和结肠弯曲菌的基因组序列长度和 GC 含量信息。利用 Python[15] 的 Matplotlib 模块将以上信息进行可视化。

1.3　基因注释信息分析

为了对弯曲菌全基因组的基因信息进行识别和标记,采用 Prokka 软件[16] 对空肠弯曲菌和结肠弯曲菌进行基因注释,后通过自编脚本从注释结果中提取密码子序列(Codon Sequence,CDS)、转运 RNA(tRNA)、转运-信使 RNA(tmRNA)、核糖体 RNA(rRNA)以及外显子(Exon)的数量和密度信息。

1.4　耐药基因分析

耐药性弯曲菌菌株的比例在全世界范围内正在快速增加[17]。通过在线软件 ResFinder 3.2[18] (https://cge.cbs.dtu.dk/services/ResFinder/)查找空肠弯曲菌和结肠弯曲菌基因组中的耐药基因并进行比较。ResFinder 采用默认参数,设置%ID 的阈值为 90%,最小长度比例为 60%。

1.5　多位点序列分型分析

MLST 方法通过测定多个管家基因(housekeeping genes)的核苷酸序列的变异情况对菌株类型进行表征。利用在线软件 MLST 2.0 (https://cge.cbs.dtu.dk/services/MLST/)[19] 对全部弯曲菌菌株进行多位点序列分型,接着利用自编脚本提取每一株弯曲菌的等位基因谱,然后分析空肠弯曲菌和结肠弯曲菌之间等位基因谱的差异。

1.6　CRISPR-Cas 系统分析

为了分析空肠弯曲菌和结肠弯曲菌在 CRISPR-Cas 系统方面的差异,利用在线软件 CRISPRCasFinder(https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/)[20] 分别对空肠弯曲菌和结肠弯曲菌中的CRISPR-Cas系统进行分析,并利用自编脚本从 CRISPR-Cas 系统中提取重复序列。根据 Clustal X 2.1 软件[21] 对重复序列的比对结果,将菌株间重复的重复序列剔除,得到无冗余的重复序列。再利用 Blast 软件[22] 将每一条无冗余的重复序列分别与空肠弯曲菌和结肠弯曲菌全基因组序列进行比对,以获得能够标识空肠弯曲菌/结肠弯曲菌的特异性序列。

2　结果与分析

2.1　全基因组序列分析

全基因组序列分析结果表明,结肠弯曲菌的基因组序列长度(均值:172.09 kbp)显著大于空肠弯曲菌(均值:166.29 kbp)(Wilcoxon 秩和检验[23],p-value = 0.000 033),如图 1a 箱线图[24] 所示。图中,绿色三角形代表均值,红色横线代表中位数。另外,结肠弯曲菌的 GC 含量(均值:31.38%)也显著高于空肠弯曲菌(均值:30.43%)(Wilcoxon 秩和检验,p-value = 0),见图 1b。

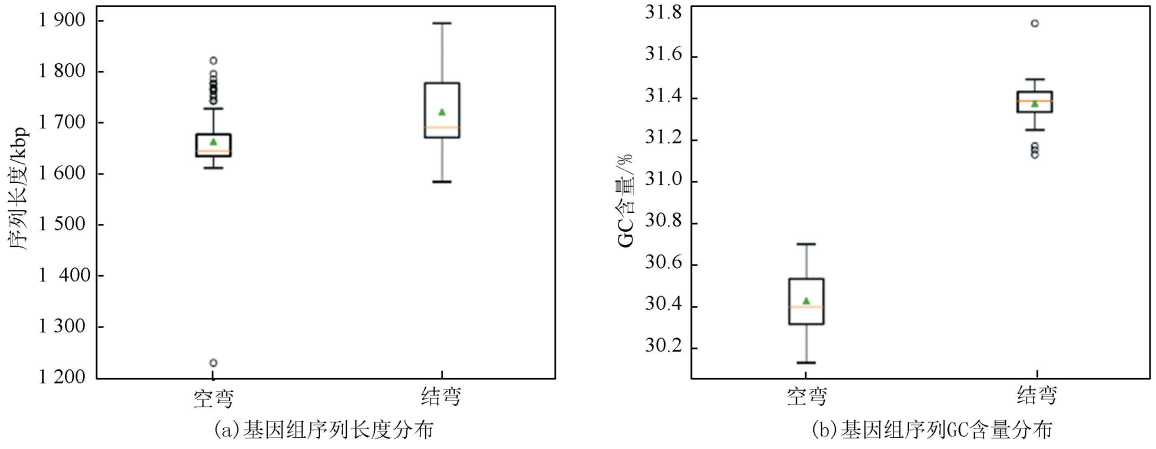

图 1 空肠弯曲菌和结肠弯曲菌的基因组序列特征比较

Fig.1 Feature comparison of genome sequences of *Campylobacter jejuni* and *Campylobacter coli*

2.2 基因组注释结果分析

利用 Prokka 软件对弯曲菌全基因组进行基因注释后,分别得到 120 份空肠弯曲菌和 22 份结肠弯曲菌的基因注释文件(GFF 格式)。通过自编脚本从注释结果中统计出每个菌株的 CDS、exon、rRNA、tRNA 以及 tmRNA 的数量。其中,每一株空肠弯曲菌/结肠弯曲菌仅含有 1 个 tmRNA。

统计分析结果表明,空肠弯曲菌和结肠弯曲菌在 CDS 数量方面没有显著差异(空肠均值:1 696.99,结肠均值1 728.18;Wilcoxon 秩和检验,p-value =0.215 891),但在 exon 数量(空肠均值:51.92,结肠均值 54.27;Wilcoxon 秩和检验,p-value = 0.021 432)、rRNA 数量(空肠均值:6.96,结肠均值 8.27;Wilcoxon 秩和检验,p-value = 0.014 076)以及 tRNA 数量(空肠均值:41.96,结肠均值 43.09;Wilcoxon 秩和检验,p-value = 0.033 78)方面,空肠弯曲菌和结肠弯曲菌之间存在显著差异(见图 2)。

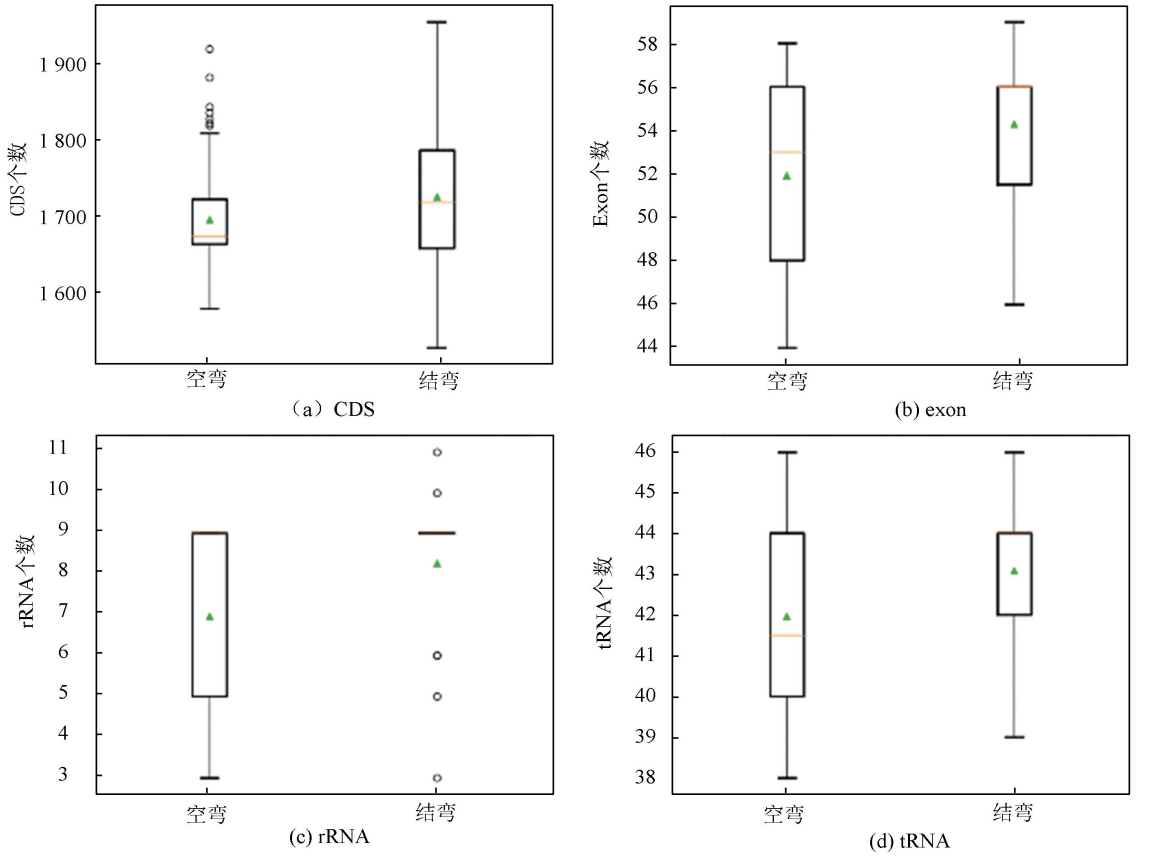

图 2 空肠弯曲菌和结肠弯曲菌的基因注释特征(绝对数量)的比较

Fig.2 Feature comparison of gene annotation (absolute count) of *Campylobacter jejuni* and *Campylobacter coli*

由于结肠弯曲菌和空肠弯曲菌在基因组序列长度方面存在显著差异,本文又对两种弯曲菌基因组上 CDS、exon、rRNA 及 tRNA 的数量密度(个数/序列长度)进行了比较(见图 3)。统计分析结果表明,空肠弯曲菌的 CDS 密度(均值:1.021 个/千碱基对)显著大于结弯肠弯曲菌(均值:1.003 个/千碱基对)(Wilcoxon 秩和检验,p-value = 0.000 37),而两种菌在 exon、rRNA 及 tRNA 的密度方面没有显著差异。

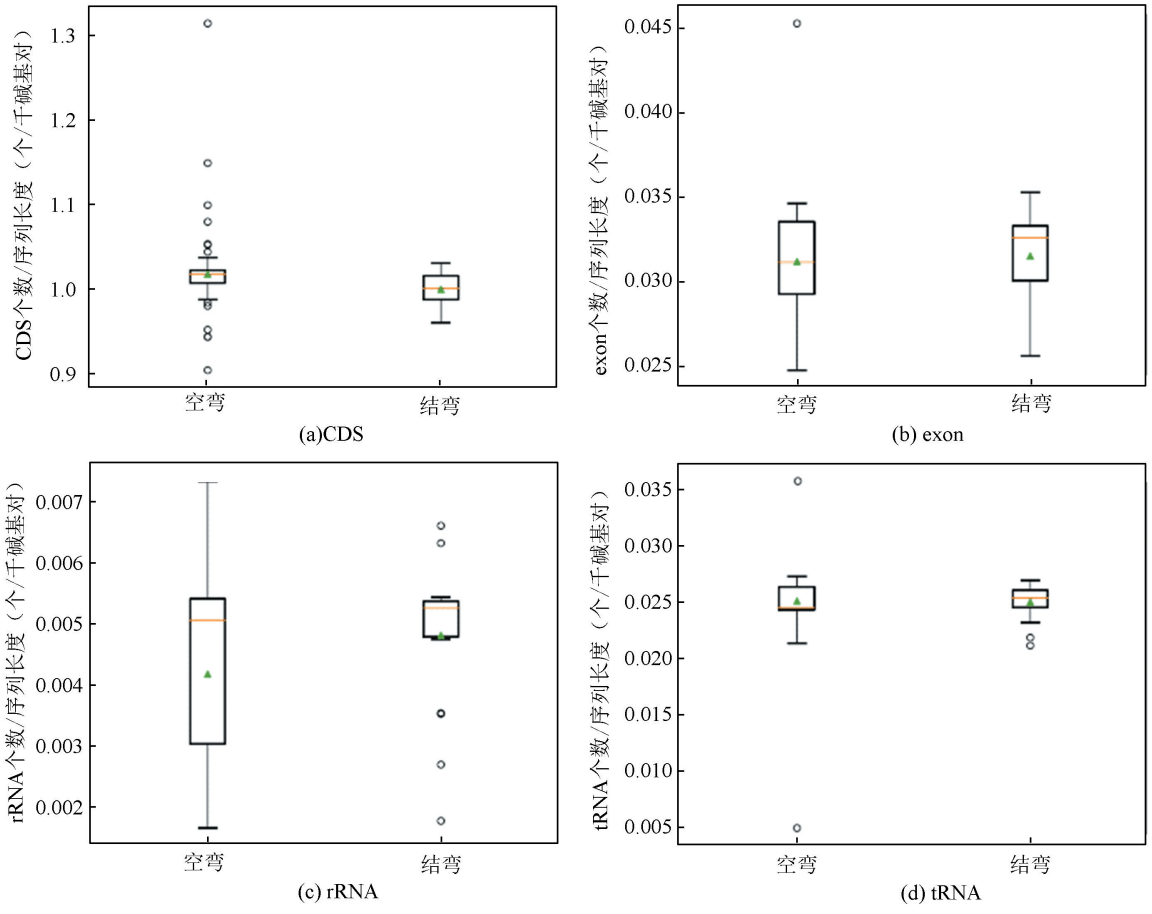

图 3　空肠弯曲菌和结肠弯曲菌的基因注释特征(数量密度)的比较

Fig.3　Feature comparison of gene annotation (count density) of *Campylobacter jejuni* and *Campylobacter coli*

2.3　耐药基因分析结果

经 ResFinder 分析发现,142 株弯曲菌共含有 $bla_{\text{OXA-193}}$、$bla_{\text{OXA-450}}$、$bla_{\text{OXA-451}}$、$bla_{\text{OXA-452}}$、$bla_{\text{OXA-453}}$、$bla_{\text{OXA-489}}$、$bla_{\text{OXA-61}}$、$bla_{\text{OXA-460}}$、$bla_{\text{OXA-465}}$、$bla_{\text{OXA-447}}$、$bla_{\text{OXA-184}}$、$bla_{\text{OXA-461}}$、*aph*(2″)-*lf*、*aadE-Cc*、*cat*、*tet*(*O*)等 16 种耐药基因。其中,含耐药基因的空肠弯曲菌和结肠弯曲菌的菌株数(比例)分别为 106 株(88.33%)和 15 株(68.18%),而发现最多的耐药基因为 bla_{OXA}(OXA β-lactamases, OXA β-内酰胺酶)型[25]。已有的研究表明,β-内酰胺类(Beta-lactam)抗生素对弯曲菌的功效有限,而对抗此类抗生素的耐药性可能是由弯曲菌产生的 β-内酰胺酶所介导的[17]。表 1 展示了含耐药基因的弯曲菌数量及占比。

统计分析结果表明,本文所采用的空肠弯曲菌和结肠弯曲菌菌株在耐药基因分布上没有显著差异(卡方检验,p-value=0.055 839)。尽管如此,两种弯曲菌菌株耐药基因的区别仍值得讨论。例如,空肠弯曲菌对除了 *aadE-Cc* 之外的大部分耐药基因均有包含,但没有发现任何结肠弯曲菌含有 $bla_{\text{OXA-465}}$、$bla_{\text{OXA-447}}$、$bla_{\text{OXA-184}}$、$bla_{\text{OXA-461}}$、*cat*、*tet*(*O*)等 6 种耐药基因。特别是在空肠弯曲菌中占比高达 45% 的耐药基因 $bla_{\text{OXA-447}}$在结肠弯曲菌中没有被发现,并且含有 $bla_{\text{OXA-447}}$的空肠弯曲菌不含其它任何 bla_{OXA}基因。

2.4　多位点序列分型结果

MLST 方法一般测定 6~10 个管家基因内部 400~600 bp片段的核苷酸序列(即 MLST 等位基因),并根据每个等位基因位点的序列被发现的顺序赋予一个等位基因编号,而每个菌株的等位基因及编号按照指定顺序排列后便构成其等位基因谱(Allelic profile)或序列型(Sequence type, ST)[26]。弯曲菌的 MLST 分析一般涵盖 *aspA*、*glnA*、*tkt*、*uncA*、

gltA、*glyA*、*pgm* 等 7 个等位基因。对 142 株弯曲菌的等位基因谱进行去冗余处理后,共得到 38 种空肠弯曲菌和 15 种结肠弯曲菌的等位基因谱(见表 2)。其等位基因编号的分布如图 4 所示。

表 1 含有耐药基因的空肠弯曲菌和结肠弯曲菌的菌株数量比较

Table 1 Count comparison of *Campylobacter jejuni* and *Campylobacter coli* strains containing resistance genes

耐药基因	空弯个数(比例%)	结弯个数(比例%)
$bla_{\text{OXA-193}}$	36(30.00)	12(54.55)
$bla_{\text{OXA-450}}$	36(30.00)	8(36.36)
$bla_{\text{OXA-451}}$	36(30.00)	8(36.36)
$bla_{\text{OXA-452}}$	37(30.83)	8(36.36)
$bla_{\text{OXA-453}}$	36(30.00)	8(36.36)
$bla_{\text{OXA-489}}$	36(30.00)	8(36.36)
$bla_{\text{OXA-61}}$	40(33.33)	8(36.36)
$bla_{\text{OXA-460}}$	4(3.33)	2(9.09)
$bla_{\text{OXA-465}}$	1(0.83)	0(0.00)
$bla_{\text{OXA-447}}$	54(45.00)	0(0.00)
$bla_{\text{OXA-184}}$	2(1.67)	0(0.00)
$bla_{\text{OXA-461}}$	3(2.50)	0(0.00)
aph(2″)-*lf*	1(0.83)	1(4.55)
aadE-Cc	0(0.00)	1(4.55)
cat	2(1.67)	0(0.00)
tet(*O*)	6(5.00)	0(0.00)

表 2 空肠弯曲菌和结肠弯曲菌的等位基因谱列表

Fig.2 List of allelic profiles of *Campylobacter jejuni* and *Campylobacter coli*

C	*aspA*	*glnA*	*gltA*	*glyA*	*pgm*	*tkt*	*uncA*	N	C	*aspA*	*glnA*	*gltA*	*glyA*	*pgm*	*tkt*	*uncA*	N
1	10	81	50	99	120	76	52	41	1	7	4	27	68	11	3	6	1
1	10	81	50	87	120	76	52	12	1	7	17	5	2	10	3	6	1
1	2	1	5	3	4	1	5	10	1	7	17	5	2	11	3	6	1
1	2	1	1	3	2	1	5	6	1	7	84	5	10	119	178	26	1
1	2	1	12	3	2	1	5	4	1	8	2	5	53	11	3	1	1
1	1	1	5	48	394	88	233	3	1	9	2	2	10	10	3	5	1
1	1	2	3	27	5	9	3	3	1	9	25	2	10	431	3	6	1
1	8	10	2	2	11	12	6	3	1	14	17	5	2	11	3	6	1
1	1	2	49	4	11	66	8	2	1	24	30	2	2	89	59	6	1
1	1	3	6	4	3	3	3	2	1	24	171	2	2	89	59	6	1
1	2	1	1	3	2	1	6	2	1	29	7	10	4	42	7	1	1
1	2	1	1	3	140	3	5	2	0	33	39	30	78	104	43	17	4
1	4	7	10	4	1	7	1	2	0	33	39	30	82	113	43	17	4
1	4	7	10	4	42	7	1	2	0	33	39	30	82	113	47	17	2
1	8	2	5	53	11	3	105	2	0	32	38	30	82	152	35	17	1
1	10	27	16	19	10	5	7	2	0	33	38	30	78	104	43	17	1
1	1	2	5	2	2	3	6	1	0	33	39	30	79	112	47	17	1
1	2	1	2	3	2	1	5	1	0	33	39	30	82	113	43	41	1
1	2	1	12	3	497	1	5	1	0	33	39	30	82	113	85	17	1
1	2	1	42	3	148	1	5	1	0	33	39	30	140	104	43	41	1
1	2	4	1	2	7	1	5	1	0	33	39	30	140	113	43	41	1
1	2	17	2	3	2	1	5	1	0	33	39	44	82	104	44	36	1
1	2	21	5	2	59	1	5	1	0	33	39	103	82	113	43	17	1
1	2	222	29	250	303	25	35	1	0	33	39	122	140	113	43	17	1
1	4	7	0	4	1	0	1	1	0	103	110	30	162	188	164	99	1
1	4	7	10	4	42	51	1	1	0	121	278	328	431	552	452	154	1
1	6	4	5	2	11	1	5	1									

注:C 列显示弯曲菌类别(1 表示空肠弯曲菌,0 表示结肠弯曲菌);N 列显示菌株数。

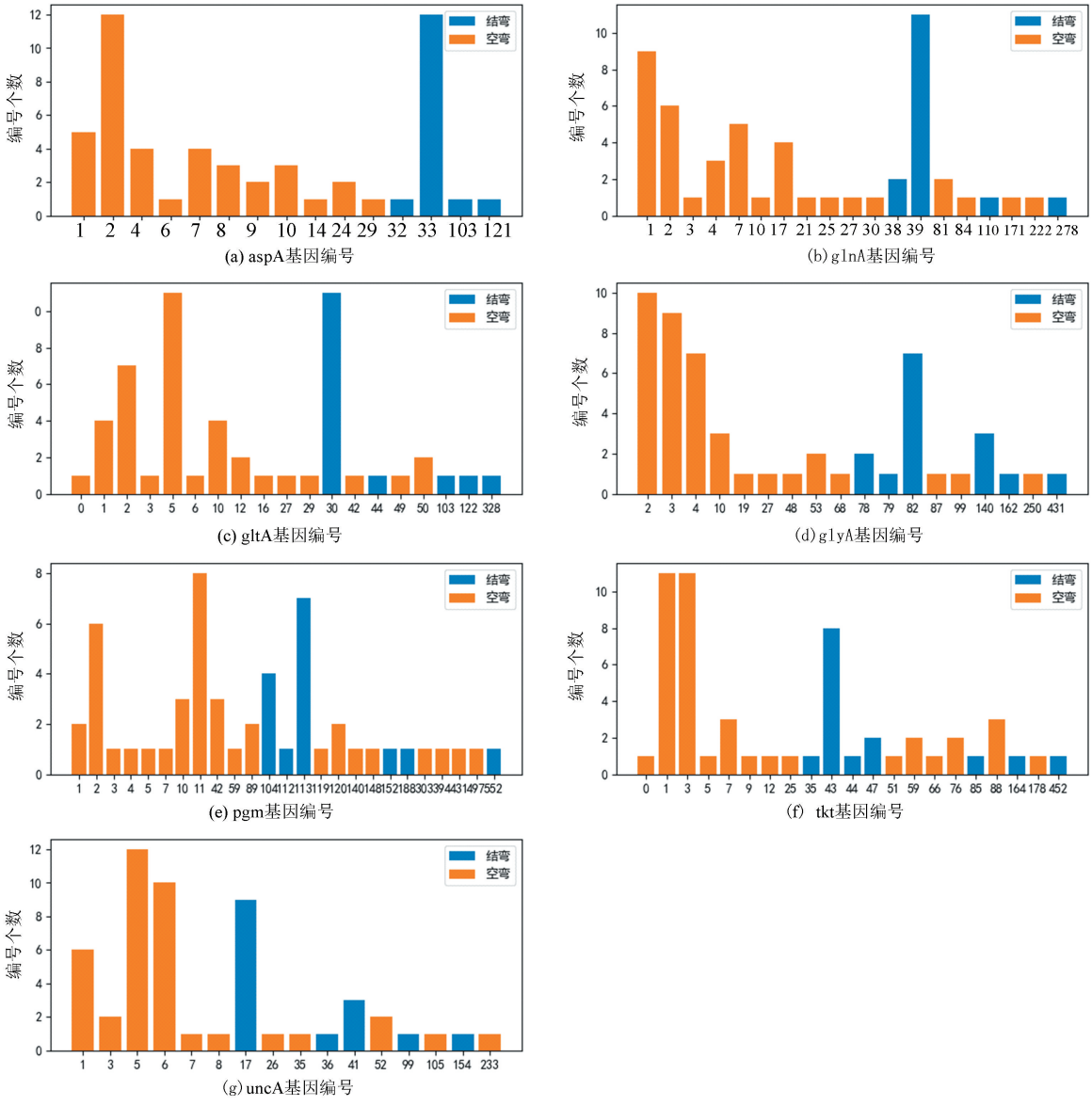

图 4　空肠弯曲菌和结肠弯曲菌的等位基因编号分布图

Fig.4　Distribution of allelic numbers of *Campylobacter jejuni* and *Campylobacter coli*

从表 2 和图 4 可见,空肠弯曲菌和结肠弯曲菌的样本菌株之间没有任何相同的等位基因谱及编号。这表明从 MLST 角度看,两种菌分属不同的家族。另外,全部空肠弯曲菌样本所覆盖的等位基因编号范围显著大于结肠弯曲菌(Wilcoxon 秩和检验,p-value = 0.001 745)。

2.5　CRISPR-Cas 系统的分析

一个完整的 CRISPR-Cas 系统包括 CAS 基因(Cas Genes)、先导序列(Leader sequence)、重复序列(Repeats/DR)以及间隔序列(Spacers)。通过 CRISPRCasFinder 分析发现大部分空肠弯曲菌(113/120,94.17%)含有至少一个 CRISPR-Cas 系统,而结肠弯曲菌中仅 12 个菌株(12/22,54.55%)含有 CRISPR-Cas 系统(见表 3)。统计分析结果表明,空肠弯曲菌和结肠弯曲菌在 CRISPR-Cas 系统数量的分布方面具有显著差异(卡方检验,p-value = 0.000 000 225 7)。

从弯曲菌样本的 CRISPR-Cas 系统中提取重复序列(DR 序列)后,利用 ClustalX 2.1 对重复序列进行比对,再将冗余序列去除,最终得到 66 条无冗余的重复序列(见图 5)。

为了找出能够标识空肠弯曲菌/结肠弯曲菌的特异性重复序列,利用 Blast 软件将 66 条重复序列分别与空肠弯曲菌和结肠弯曲菌基因组一一进行比对。结果发现重复序列 NZ_CP017859_1(见图 6)在 105 株空肠弯曲菌中存在(105/120,87.5%),而仅在 1 株结肠弯曲菌中存在(1/22,4.5%)。

表 3 弯曲菌的 CRISPR-Cas 系统统计表

Table 3 Statistics of CRISPR-Cas systems of *Campylobacter*

CRISPR-CAS 系统数量	空肠弯曲菌数量(比例%)	结肠弯曲菌数量(比例%)
0	7(5.84)	10(45.45)
1	75(62.50)	9(40.90)
2	34(28.33)	1(4.55)
3	1(0.83)	1(4.55)
4	1(0.83)	1(4.55)
5	2(1.67)	0(0.00)
合计	120(100)	22(100)

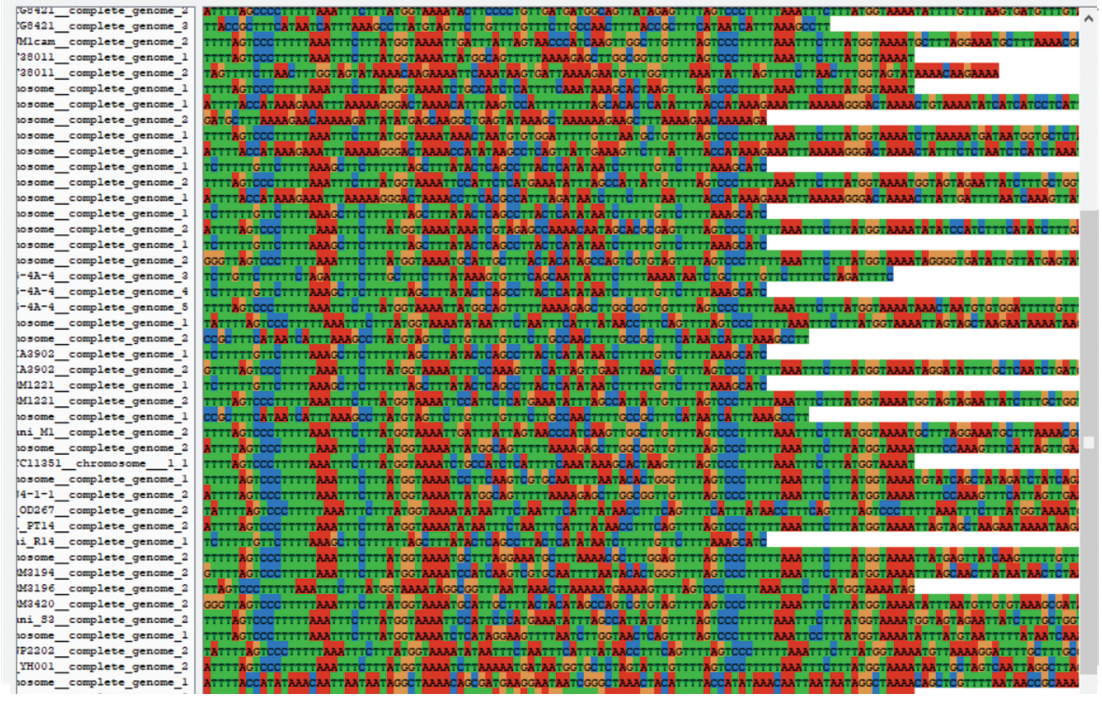

图 5 66 条无冗余重复序列(部分截图)

Fig.5 Sixty-six non-redundant repeat sequences (partial snapshot)

为了进一步验证 NZ_CP017859_1 的分辨力,本文将重复序列 NZ_CP017859_1 与江苏省人兽共患病学重点实验室的自有菌株基因组(数据尚未公开)进行 Blast 比对,结果发现自有的全部 22 株空肠弯曲菌中有 20 株含有 NZ_CP017859_1 序列(20/22,90.9%),而全部 27 株结肠弯曲菌均不含此序列。

图 6 重复序列 NZ_CP017859_1

Fig.6 A repeat sequence NZ_CP017859_1

3 结 论

利用生物信息学方法对弯曲菌全基因组进行了包括序列、基因注释、耐药基因、多位点序列分型以及 CRISPR-Cas 系统等分析。研究发现,结肠弯曲菌的基因组序列长度显著大于空肠弯曲菌,同时结肠弯曲菌的 GC 含量也显著高于空肠弯曲菌。基因

注释结果表明,每株弯曲菌仅含一个 tmRNA。结肠弯曲菌在 exon、tRNA 和 rRNA 三个特征的绝对数量上显著大于空肠弯曲菌,但两种菌的 CDS 含量接近。同时,空肠弯曲菌的 CDS 密度显著高于结肠弯曲菌,而 exon、tRNA 和 rRNA 在两种弯曲菌基因组上的密度相近。耐药基因分析结果表明,空肠弯曲菌和结肠弯曲菌在含耐药基因菌株的比例方面没有显著差异,但发现有相当一部分(45%)空肠弯曲菌含有 $bla_{\text{OXA-447}}$且不含其它 bla_{OXA}基因,可以推测这部分空肠弯曲菌菌株产生-内酰胺酶的机制会比较特殊,并且可能依赖于 $bla_{\text{OXA-447}}$。MLST 分析发现空肠弯曲菌和结肠弯曲菌有着截然不同的等位基因谱,且前者的编号范围较宽。CRISPR-Cas 系统分析表明,大部分空肠弯曲菌(94.16%)含有 CRISPR-Cas 系统,但对于结肠弯曲菌只有54.55%。另外,研究发现重复序列 NZ_CP017859_1 普遍存在于空肠弯曲菌中,而极少结肠弯曲菌含有此序列,故利用重复序列 NZ_CP017859_1 进行空肠弯曲菌鉴定具有较高的敏感性和特异性。综上,可认为序列长度、GC 含量、exon、tRNA 和 rRNA 的数量、CDS 密度、等位基因谱和 CRISPR 重复序列 NZ_CP017859_1 可以作为区分空肠弯曲菌和结肠弯曲菌的特征,其中重复序列 NZ_CP017859_1 具有较高的高分辨力。这些特征可以结合支持向量机和深度学习等机器学习/人工智能方法,以提高空肠弯曲菌和结肠弯曲菌鉴别的准确性。同时,通过增加结肠弯曲菌的样本数量,构建平衡样本集,能够进一步优化预测模型。

参考文献(References)

[1]EKDAHL K, NORMANN B, ANDERSSON Y. Could flies explain the elusive epidemiology of campylobacteriosis? [J]. BMC Infectious Diseases, 2005, 5(1):11. DOI:10.1186/1471-2334-5-11.

[2]SAHIN O, KOBALKA P, ZHANG Q. Detection and survival of Campylobacter in chicken eggs[J]. Journal of Applied Microbiology, 2003, 95(5):1070-1079. DOI: 10.1046/j.1365-2672.2003.02083.x.

[3]BOES J, NERSTING L, NIELSEN E M, et al. Prevalence and diversity of campylobacter jejuni in pig herds on farms with and without cattle or poultry[J]. Journal of Food Protection, 2005, 68(4):722-727. DOI:10.4315/0362-028X-68.4.722.

[4]CHEN Jie, SUN Xinting, ZENG Zheng, et al. Campylobacter enteritis in adult patients with acute diarrhea from 2005 to 2009 in Beijing, China[J]. Chinese Medical Journal, 2011, 124(10):1508-1512. DOI: 10.3760/cma.j.issn.0366-6999.2011.10.013.

[5]龚俊, 刘树林. 空肠弯曲菌与大肠弯曲菌基因分型研究进展[J]. 中华微生物学和免疫学杂志, 2004(05):81-85.
GONG Jun, LIU Shulin. Advances in genotyping of *Campylobacter jejuni* and *Campylobacter coli*[J]. Chinese Journal of Microbiology and Immunology, 2004(05):81-85.

[6]袁宝君. 多重 PCR 与传统方法检测空肠弯曲菌的比较研究[J]. 卫生研究,2007(01):98-100. DOI:10.3969/j.issn.1000-8020.2007.01.031.
YUAN Baojun. Comparative study of multiplex PCR and traditional methods for detecting *Campylobacter jejuni*[J]. Health Research, 2007(01):98-100. DOI:10.3969/j.issn.1000-8020.2007.01.031.

[7]FITZGERALD C, WHICHARD J, NACHAMKIN I, et al. Diagnosis and antimicrobial susceptibility of Campylobacter species[M]. 2nd ed. Campylobacter. 2nd ed. Washington D C: ASM Press, 2000: 45-66.

[8]LUND M, NORDENTOFT S, PEDERSEN K, et al. Detection of Campylobacter spp. in chicken fecal samples by real-time PCR[J]. Journal of Clinical Microbiology, 2004, 42(11):5125-5132. DOI: 10.1128/JCM.42.11.5125-5132.2004.

[9]RIDLEY A M, ALLEN V M, SHARMA M, et al. Real-time PCR approach for detection of environmental sources of Campylobacter strains colonizing broiler flocks[J]. Applied and Environmental Microbiology,2008,74(8):2492-2504. DOI:10.1128/AEM.01242-07.

[10]RÖNNER A, LINDMARK H. Quantitative detection of campylobacter jejuni on fresh chicken carcasses by real-time PCR[J]. Journal of Food Protection, 2007, 70(6):1373-1378. DOI:10.4315/0362-028X-70.6.1373.

[11]LLARENA A K, TABOADA E, ROSSI M. Whole-genome sequencing in the epidemiology of Campylobacter jejuni infections[J]. Journal of Clinical Microbiology, 2017, 55(5):1269-1275. DOI:10.1128/JCM.00017-17.

[12]莫惠栋, 顾世梁. 基因组长度的估计方法[J]. 科学通报,2000, 45(13): 1414-1419. DOI: 10.3321/j.issn:0023-074X.2000.13.013.
MO Huidong, GU Shiliang. The estimation method of genome length[J]. Chinese Science Bulletin,2000,45(13):1414-1419. DOI: 10.3321/j.issn:0023-074X.2000.13.013.

[13]王芳平, 王志坚, 李永香. 三种模式微生物基因组中 GC 含量的比较[J]. 基因组学与应用生物学,2019(5):2215-2220.
WANG Fangping, WANG Zhijian, LI Yongxiang. Comparison of GC contents in three microbial genomes[J]. Genomics and Applied Biology, 2019(5):2215-2220.

[14]COCK P J A, TIAGO A, CHANG J T, et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics[J]. Bioinformatics, 2009(11):

1422-1423. DOI:10.1093/bioinformatics/btp163.

[15] OLIPHANT T E. Python for scientific computing[J]. Computing in Science & Engineering,2007,9(3):10-20. DOI: 10.1109/MCSE.2007.58.

[16] SEEMANN T. Prokka: Rapid prokaryotic genome annotation[J]. Bioinformatics,2014,30(14):2068-2069. DOI: 10.1093/bioinformatics/btu153.

[17] LUANGTONGKUM T, JEON B, HAN J, et al. Antibiotic resistance in Campylobacter: Emergence, transmission and persistence[J]. Future Microbiology, 2009,4(2):189-200. DOI:10.2217/17460913.4.2.189.

[18] ZANKARI E, HASMAN H,COSENTINO S,et al. Identification of acquired antimicrobial resistance genes[J]. J Antimicrob Chemother,2012,67: 2640-2644. DOI:10.1093/jac/dks261.

[19] AHMED A, FERREIRA A S, HARTSKEERL R A. Multilocus sequence typing (MLST): Markers for the traceability of pathogenic Leptospira strains[J]. Methods in Molecular Biology,2015,1247:349-359. DOI:10.1007/978-1-4939-2004-4_25.

[20] IBTISSEM G, GILLES V, CHRISTINE P. CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats[J]. Nucleic Acids Research, 2007,35(Web Server issue): W52 - W527. DOI: 10. 1093/nar/gkm360.

[21] LARKIN M A, BLACKSHIELDS G, BROWN N P, et al. Clustal W and Clustal X version 2.0[J]. Bioinformatics, 2007, 23(21):2947-2948. DOI:10.1093/bioinformatics/btm404.

[22] SCOTT M G, MADDEN T L. BLAST: At the core of a powerful and diverse set of sequence analysis tools[J]. Nucleic Acids Research, 2004,32(Web Server issue): W20-W25. DOI:10.1093/nar/gkh435.

[23] HAYNES W. Wilcoxon Rank Sum Test [A]. DUBITZKY W, WOLKENHAUER O, CHO K, et al. Encyclopedia of Systems Biology. New York, NY: Springer, 2013: 2354-2355. DOI:10.1007/978-1-4419-9863-7_1185.

[24] SCHWERTMAN N C, OWENS M A, ADNAN R. A simple more general boxplot method for identifying outliers [J]. Computational Statistics & Data Analysis, 2004, 47(1):165-174. DOI:10.1016/j.csda.2003.10.012.

[25] EVANS B A, AMYES S G B. OXA β-Lactamases[J]. Clinical Microbiology Reviews, 2014, 27(2):241. DOI: 10.1128/CMR.00117-13.

[26] MAIDEN M C J. Multilocus sequence typing of bacteria [J]. Annual Review of Microbiology, 2006, 60(1): 561-588. DOI:10.1146/annurev.micro.59.030804.121325.

[责任编辑:吴永英]